DOT/FAA/AM-95/32

Office of Aviation Medicine
Washington, D.C. 20591

Some Effects of 8- vs. 10-Hour Work Schedules on the Test Performance/Alertness of Air Traffic Control Specialists

David J. Schroeder
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73132

Roger R. Rosa
National Institute for Occupational Safety and Health
Cincinnati, Ohio

L. Alan Witt
Barnett Banks
Jacksonville, Florida

December 1995

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-95/32	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Some Effects of 8- vs. 10-Hour Work Schedules on the Test Performance/Alertness of Air Traffic Control Specialists		5. Report Date December 1995 6. Performing Organization Code
7. Author(s) David J. Schroeder, Ph.D. FAA Civil Aeromedical Institute Roger R. Rosa, Ph.D. National Institute for Occupational Safety and Health Division of Biomedical and Behavioral Science L. Alan Witt, Ph.D. Barnett Banks		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplemental Notes		

16. Abstract

A 10-hour, 4-day rotating shift schedule worked by some Air Traffic Control Specialists (ATCSs) was compared to the more traditional 8-hour, 2-2-1 rapidly rotating schedule. Measures of performance and alertness were obtained from a group of 52 ATCSs at an en route ATC center on tasks in the NIOSH fatigue test battery. Additional information on sleep patterns, mood, and somatic complaints was also gathered. Results confirm that tests comprising the NIOSH test battery are sensitive to fatigue and diurnal variations associated with a rotating shift schedule. Test performance of ATCSs on the 10-hour shift did not differ from those on the 8-hour schedule for any of the NIOSH parameters, when comparing the initial 4 days of the work week. Test performance was notably poorer on the mid-shift (night) that occurred on the final (fifth) day of the 2-2-1 8-hour schedule. For both schedules, there was evidence of changes in alertness on some of the NIOSH performance measures within work days and across days of the week. Changes in test performance and mood ratings corresponded to the decline in self-reported sleep time across the work week.

17. Key Words Shift Work, Mood, Fatigue, Sleep, Human Performance, Air Traffic Control		18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 20	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SOME EFFECTS OF 8- VS. 10-HOUR WORK SCHEDULES ON THE TEST PERFORMANCE/ALERTNESS OF AIR TRAFFIC CONTROL SPECIALISTS

INTRODUCTION

During the past 10 years, management has been faced with increased employee demands for more flexible work schedules, including interest in "compressed" work schedules. A "compressed" work schedule refers to any work week where employees are allowed to complete their work in 4 or fewer days. Numerous questions have been raised concerning the possible effects of compressed schedules on productivity, job efficiency and fatigue, and associated concerns with safety and health.

Empirical research on the impact of compressed work schedules has focused more closely on employees' subjective reports concerning fatigue, alertness, mood, job satisfaction and conflicts with family activities, and leisure time. Outcomes have indicated: (a) increases in organizational effectiveness (Hartman & Weaver, 1977; Wheeler, 1970), as well as no increases (Calvasina & Boxx, 1975); (b) increased satisfaction brought about by more leisure time (Hodge & Tellier, 1975; Steele & Poor, 1970) but greater fatigue, conflict with evening activities, and conflict between the work schedule and family and child-related activities (Hodge & Tellier, 1975; Kenny, 1974); (c) a full range of positive, negative, and neutral affective responses (cf. Dunham, Pierce, & Castaneda, 1987); and (d) both greater fatigue (e.g., Goodale & Aagaard, 1974; Hodge & Tellier, 1975) and no differences in fatigue (Latack & Foster, 1985).

Changes in performance and alertness associated with compressed work schedules have, until recently, received less attention. Of the performance-based studies, nearly all involve comparisons of 8-h and 12-h shift schedules. In an early exception, Volle, et al. (1979), reported that factory employees on the 10-h versus 8-h schedule did not differ significantly on reaction time but did display decreased grip strength and higher critical flicker fusion (CFF) thresholds. However, the authors concluded that the increase in fatigue remained within acceptable limits and that there was no evidence that these changes affected overall productivity in the manufacturing plant. Peacock, et al. (1983) on the other hand, found in a study of police officers, improved subjective alertness, sleep, and cardiovascular fitness (12-h versus 8-h). No significant differences were noted on CFF thresholds or grammatical reasoning tests. Mills, Arnold & Wood (1983), while not making a direct comparison with nurses on an 8-h schedule, noted that employees on a 12-h shift schedule evidenced significant increases in subjective fatigue and grammatical reasoning errors from start to completion of the workday. However, a majority of the increase in errors occurred between the 1st and 6th hours of the workday. The nurses did perform more rapidly on the grammatical reasoning test across the workday and expressed high levels of satisfaction with the 12-h schedule. Daniel & Potasova (1989) also reported some differences between 12-h and 8-h personnel on several cognitive and psychomotor tasks; however, these findings may have been influenced by differences in initial performance capabilities between the 2 groups. Lewis & Swaim (1988), utilizing a number of measures of employee performance and fatigue, compared the effects of 8-h and 12-h shift schedules at an experimental nuclear reactor. While the results were mixed, with some indications of greater fatigue on the 12-h schedule, direct on-the-job performance measures favored the 12-h shift. A vast majority of the employees favored the 12-h schedule and the authors concluded that the 12-h shift schedule was a "reasonable alternative to an 8-h schedule (p. 513)."

The computerized National Institute of Occupational Safety and Health (NIOSH) Fatigue Test Battery was developed to quantify changes in several indices of cognitive, sensory, and perceptual-motor performance and self-reported subjective feelings associated with shift work. As part of that development, Rosa, et al. (1985) assessed differences in the test performance of subjects working 6 8-h days or 4

A-1

12-h days. They found that individuals on the 12-h 4-day work week reported greater fatigue on several of the self-report measures than when on an 8-h 6-day work week. Greater evidence of fatigue was also found on the grammatical reasoning and digit addition cognitive performance measures from the test battery. In a subsequent laboratory investigation to assess the effects of fatigue and diurnal variations on performance on the test battery, Rosa & Colligan (1988) compared the effects of working 5 12-h days in the laboratory, using a data entry job simulation task with rest periods. They found that changes in performance on the data entry task associated with the workday and work week corresponded closely with subjective ratings and performance on a number of the tasks comprising the test battery. Rosa & Colligan (1988) concluded that the NIOSH Fatigue Test Battery is sensitive to long hours of work and to the influence of circadian rhythms on performance. In field studies, Rosa, Colligan & Lewis (1989) and Rosa & Bonnet (1993) found evidence of significant differences in self-reported sleep time and fatigue, as well as performance on some aspects of the test battery, when comparing employees who were working 8-h and 12-h shift schedules at gas utilities and continuous processing plants. Thus, the findings of Rosa and his colleagues confirm that employees working a 12-h compressed work schedule experience greater fatigue and exhibit lower performance capabilities on some test measures as compared to those on more traditional 8-h schedules. In a 3 to 5 year follow-up, Rosa (1991) found that the sleep loss and performance declines were still present in employees on the 12-h shift schedule. However, employees still expressed generally high levels of satisfaction concerning the 12-h shifts and there was no operational evidence that safety was compromised by the associated fatigue. The lack of any demonstrable change in the operational performance measures may be due to the fact that the performance measures are not sufficiently rigorous to detect the effects of fatigue or that the performance requirements in the operational environment do not require as quick performance as is measured in the various NIOSH fatigue tests.

The FAA has approved the use of compressed and flexible work schedules for its employees, including air traffic control specialists (ATCSs). This action included temporary approval for the use of 10-h workdays. While FAA management closely reviewed various ATCS performance parameters to identify possible negative affects from the 10-h schedule, they also decided that a scientific study should be undertaken to assess the potential effects of working the 4-day 10-h shift schedule on employee performance capabilities. Since there is little information available in the literature concerning the 10-h workday, and the ATC work environment is sufficiently unique from the work environments included in the above-mentioned studies, this study was initiated to compare the effects of the existing 2-2-1 8-h shift schedule with that of the 4-day 10-h schedule on measures of employee cognitive performance and self-reported sleep and mood.

METHOD

Measurements

NIOSH Fatigue Test Battery. This flexible, computerized test battery was developed by Rosa, et al. (1985) and Rosa & Colligan (1988) specifically for applications in field experimentation with employees working on different shift schedules. Users can select from a group of tests that assess cognitive, perceptual-motor, and motor skills. Additional tests and self-report measures of alertness, fatigue, and the quality and duration of sleep can be incorporated into the battery, with limited programming requirements. Flexibility is also provided by the ability to tailor the test length to the research requirements and available time. The investigator is thus able to construct a test battery that is highly responsive to the job demands and requirements of the work setting. The choice reaction time, mental arithmetic, and grammatical reasoning tests were selected for inclusion in this study both on the basis of their demonstrated sensitivity to alterations in alertness and association with the job tasks of an ATCS. The relevance of these tasks to the ATC occupation is further supported by recent findings of Broach and Aul (1993), who used interviews of ATCSs and subsequent ratings on the Position Analysis Questionnaire to identify attributes of abilities or aptitudes required of ATCSs. Of greater relevance were perceptual speed, closure, reaction

time, and short-term memory. Numerical computation, arithmetic reasoning, and convergent and divergent thinking were also somewhat more relevant for the ATC profession than for other jobs.

The choice reaction time task consisted of random presentation of the words TRUE or FALSE on the VDT for a total of 150 trials over approximately 10-m. The intertrial interval was random, with a range of 2 to 5-s. Subjects were required to press a push-button switch labeled "TRUE" or "FALSE" on a specially developed response box as quickly as possible to indicate the correct word. The ATCS's forefinger and middle finger of his/her preferred hand rested on the buttons during the trial. For this study, scores for the CRT task included the mean reaction time and number of errors (i.e., incorrect responses).

The mental arithmetic test is an adaptation of the test developed by Williams & Lubin (1966). At the beginning of the task, a randomly selected constant between the values of 3 through 9 was presented for 3-s and then removed for the remainder of the task. ATCSs were required to add the constant to the sum of 2 single digits and then type the last digit of the overall sum on the keyboard. The digits varied across trials and were generated immediately after an ATCS's response. Scores for the task included the number correct and number of errors during the 3-m time period.

The grammatical reasoning task is a variation of the well-known task first devised by Baddeley (1968). In this 16-trial task, a 3-letter stimulus string (e.g., JLN) was presented for 2-s, removed, and then followed after 3-s with a conditional statement such as "**J DOES NOT PRECEDE N.**" The ATCS was required to press a push-button switch labeled "TRUE" or "FALSE" as quickly as possible to indicate whether the statement described the letter string. Scores for the GR task included average response latency for correct responses and total number of errors.

At the beginning of each testing session, subjects responded to 10 choice reaction time trials and 60-s of digit addition. These mini-sessions served the dual purpose of providing a "warm-up" and resolving any potential software or hardware problems before commencing the full battery.

Daily sleep, somatic complaints, and mood. The test battery was programmed to include items about amount of sleep, ratings of quality of sleep, mood, and somatic complaints. Subjects were asked to indicate their time of retiring, arising, sleep latency, and number of awakenings. A series of 5 questions, utilizing 5-point Likert rating scales, were included to evaluate the depth and quality of sleep. Workers also responded to the 29 item (19 positive and 10 negative) Naval Psychiatric Research Unit (NPRU) mood scale (Johnson & Naitoh, 1974). ATCSs also provided ratings of workload, using the Task Load Index (TLX) scale developed by the National Aeronautics and Space Administration (NASA). Results of the workload data were not included in these analyses. Respondents were asked to indicate the presence or absence of each of 19 potential somatic complaints (e.g., headache, back pain, etc.).

Subjects

A total of 56 ATCSs (mean age = 37.9 years) from an en route air traffic control center initially volunteered to participate in the study. Prior to the initiation of the study, each ATCS was provided a description of the proposed study and asked to sign a consent form concerning the research project. A numerical code was assigned to each subject for the test sessions to ensure anonymity. Of this group of 56, 26 ATCSs working the 8-h 2-2-1 schedule and 26 on the 10-h 4-day schedule completed a sufficient number of sessions (10 or more) to be included in the study.

Work Schedules

The 8-h and 10-h rotating shift schedules are illustrated in Figure 1. Under the 8-h rapidly rotating, phase advancing schedule, ATCSs worked 2 consecutive afternoons, 2 mornings, and then returned on a quick turnaround to work an evening shift. This schedule has been in use in ATC facilities for many years, and a considerable body of research in the 1970s was dedicated to evaluating the 2-2-1 8-h schedule versus a straight 5-day 8-h rotating shift schedule (Melton, et al., 1971; 1973; 1975; and Saldivar, Hoffman, & Melton, 1977). On the 2-2-1 schedule, there are 2 nights when the time between the end of one shift and the beginning of another is sufficiently short to reduce the amount of available sleep time. The average number of hours between the end of the workday on day 1 and the start of the workday on day 2, and between day 3 and 4 was approximately 15; the average was approximately 9.3 hours between day 2 and 3, and 8.1 hours between day 4 and 5. ATCSs on the 10-h schedule also had variable starting times across the 4 days; working 2 afternoons followed by 2

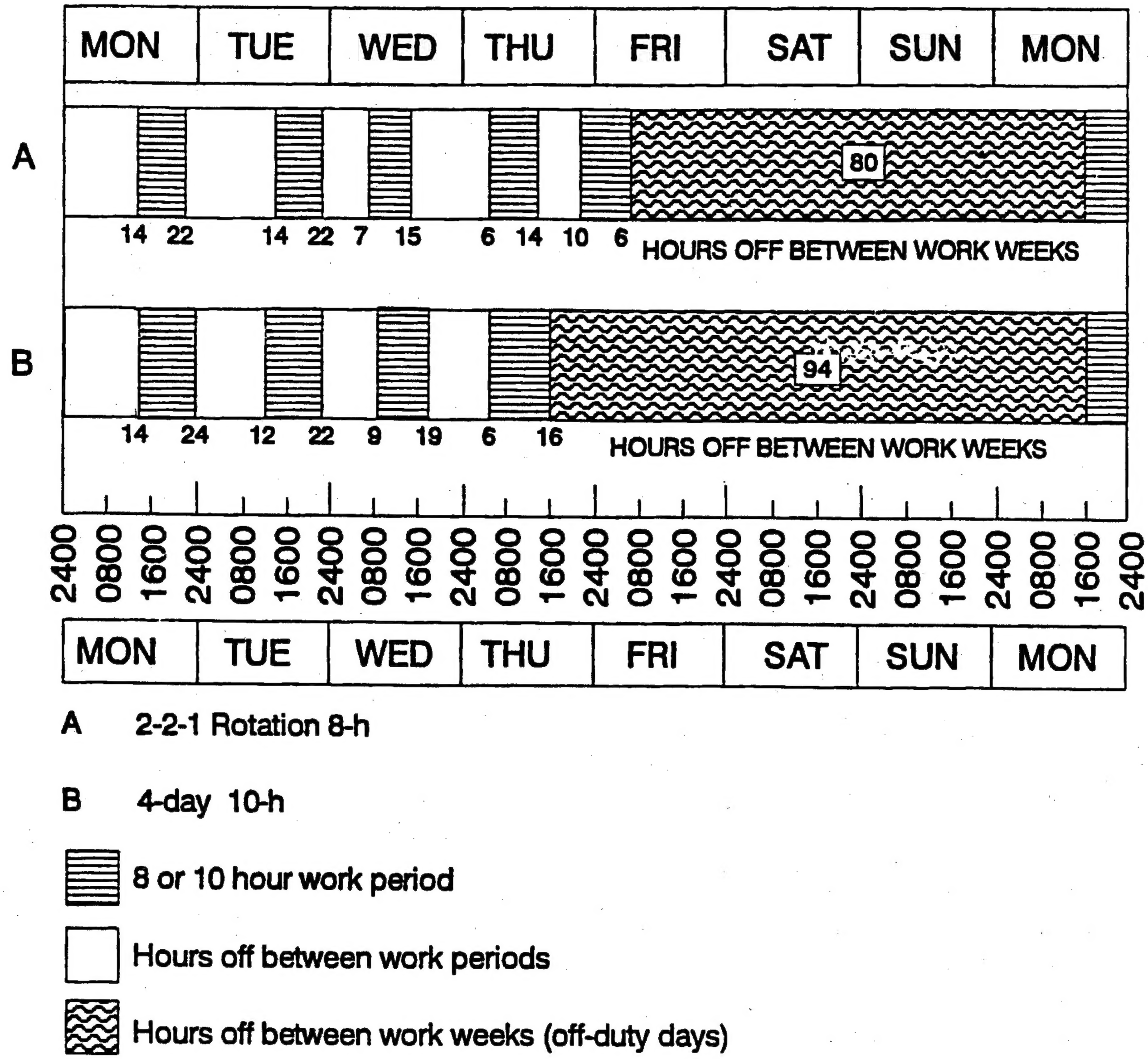


Figure 1. A graphic representation of a work week on the 2-2-1 8-h and 4-day 10-h shift schedules. The weeks shown begin on Monday; however, an ATCS's work week may start on any day and the week end will come on days other than Saturday and Sunday.

mornings. On the 10-h schedule, the average number of hours between the end of a workday and the start of the next was: 12.2 between day 1 and 2; 11.1 between day 2 and 3; and 11.2 between day 3 and 4.

Test Schedule

The NIOSH Fatigue Test Battery was administered on 3 occasions during the course of each workday. The initial session was conducted at the time the ATCS arrived at the facility. Session 2 was completed 2 hours prior to the end of the workday (at the end of 6-h for the 8-h participants and 8-h for the 10-h participants). The third and final session was administered at the close of the workday. To the extent possible, the test sessions for each ATCS were administered at the same time across each of the 3 weeks. Some disruption in the test schedule occurred for 1 of the groups of 2-2-1 8-h and 10-h subjects as a result of a snow storm, which restricted travel to and from the ATC en route facility for several days.

DATA COLLECTION PROTOCOL

O HOURS	6/8 HOUR	8/10 HOURS
Testing Time 35 min	**Testing Time 30 min**	**Testing Time 30 min**
Sleep amount	----	----
Sleep quality	---	---
Current mood	**Current mood**	**Current mood**
Physical complaints	**Physical complaints**	**Physical complaints**
TLX workload	**TLX workload**	**TLX workload**
Reaction time	**Reaction time**	**Reaction time**
Arithmetic	**Arithmetic**	**Arithmetic**
Grammatical reasoning	**Grammatical reasoning**	**Grammatical reasoning**

Figure 2. Approximate test times for each of the 3 test sessions and the information and performance data gathered during each of those sessions.

Procedure

Seven microcomputers for administering the NIOSH fatigue test battery were located in a separate room within the en route center. The computers contained all of the instructions for completing each test session, along with the performance tests and rating scales. All response data were also stored on the computer. Following introduction to the computers and the test battery, test sessions were self-administered. An experimenter was available throughout the testing period to respond to questions and to intervene if a problem occurred with the computer. Figure 2 provides an overview of the test times and test elements included during each of the 3 sessions during a workday.

Test Battery Data Analysis

Data exclusions. Data from overtime days (only 5 or 6 days total), and dubious performance sessions were excluded from statistical analyses. Choice reaction time scores were excluded if more than 50 errors occurred. Digit addition scores were excluded if more than 45 errors occurred. Grammatical reasoning scores were excluded if more than 7 errors (i.e., 50% or more errors) occurred. These exclusion criteria are consistent with procedures used in previous studies associated with the NIOSH Fatigue Test Battery.

There were only 12 problematic scores for the choice reaction time measure. For digit addition errors, from 0-14% of the responses were excluded. There was no evidence that the problematic responses for those measures were systematically related to time of day or day of the week. A slightly higher proportion of the grammatical reasoning scores were excluded on the basis of a high number of errors (0-18%) or unusually fast response times (4 - 21%). For the latter measure, there was a significant interaction between day and session, based on an ANOVA calculated only on scores from the 8-hour group ($F(2,192) = 2.78$, $p<.01$). The higher percentage of problematic scores on the grammatical reasoning test is consistent with other outcomes. In general, the ATCSs who participated in this study performed more accurately on the computer-based tasks than did other work groups (clerical/office personnel, control room operators, and gas control workers, Rosa and Colligan, 1988; Rosa and Bonnet, 1993; and Rosa, Colligan, and Lewis, 1989).

Data transformations. The procedures for transforming the data were consistent with those used in previous studies and are based on the recommendations of Myers (1979). Several of the dependent variables in the test battery were transformed to approximate a normal distribution. Grammatical reasoning response time and choice reaction

time were transformed to their inverses. Grammatical reasoning errors, choice reaction time errors, and digit addition errors were analyzed as percentage scores transformed to the arcsine of their square roots.

Analysis of variance. The effects of shift schedule, workdays, test sessions within workdays, and their interactions were tested for statistical significance with analysis of variance (ANOVA). For these between shift comparisons, day 5 of the 8-h shift schedule (night shift) was excluded from the ANOVAs. Because of unequal cell frequencies, least-squares regression solutions to the ANOVAs were computed using the SAS General Linear Models Procedure. In addition, supplementary repeated-measures ANOVAs were performed within each schedule (i.e., excluding any data from the other shift schedule) testing the effects of workdays, sessions, and their interaction. These supplementary ANOVAs eliminated between-group variance to obtain a more powerful test of changes within a shift schedule. Day 5 of the 8-h schedule was included in the within-schedule ANOVAs for that schedule. An alpha level of $p<0.05$ was considered statistically significant in all analyses.

RESULTS

NIOSH Fatigue Test Battery

Table 1 lists significant effects from the between shift ANOVAs along with a brief description of each effect. Table 2 provides a listing of the ANOVAs and a brief description of each effect for the within shift comparisons. Of the various comparisons, there were no instances where differences in NIOSH test performance between ATCSs on the 8-h and 10-h shifts were statistically significant. Performance related differences were generally due to effects associated with day of the work week, sessions, and the day-by-sessions interactions.

Effects of the work schedule on choice reaction time (RT) performance (mean reaction time and errors) are presented in Figure 3. ATCSs on both the 8-h and 10-h schedules exhibited a steady increase in reaction times from the initial to final workday during their respective work weeks, with the 10-h group exhibiting slightly quicker overall average reaction times. The slowest reaction time occurred for ATCSs during the mid-shift, where the mean reaction time (.482 seconds) was approximately 12% above the average noted on the first day of the work week. With the exception of day 4 and the mid-shift for the 8-h group, the average reaction time for the final session of the workday was consistently faster than that of the first or second session. While the number of errors associated with the choice reaction time task did not increase across the work week, the between shifts comparisons yielded significant effect for session and the day-by-session interaction. The average number of errors generally increased from the first to final session within a workday. The increase was more pronounced on the third, fourth, and fifth days of the work week.

Effects of the 2 shift schedules on ATCS performance on the digit addition task (number attempted and number of errors) are presented in Figure 4. Across days of the work week, there was a general increase in number of problems attempted. However, for the 8-h schedule, the number attempted only increased through day 3. On days 4 and 5, ATCSs on the 8-h schedule completed fewer problems than on either of the 3 previous days. While the between shift comparison revealed a significant day effect for the number of errors measure, none of the subsequent within shift comparisons reached statistical significance. The average number of errors remained relatively stable across the first 4 days of the work week for both the 8-h and 10-h groups, the average for the 8-h group on day 5 was clearly above that of any of the preceding days.

Performance on the grammatical reasoning test (response time and errors) is presented in Figure 5. The significant between shift day effect is associated with the general decline in response time for ATCSs in both groups from the first day through day 3 of the work week. On the fourth day, ATCSs on the 10-h schedule exhibited response times that were quicker than those on day 3, while ATCSs on the 8-h shift had slightly slower response times than on day 3. On the fifth day (mid shift), the average response times of ATCSs on the 8-h shift were comparable to those noted on the second and fourth days. ATCSs in both groups had quicker response times during the final session than on the other 2 sessions of each day (with one exception).

TABLE 1

BETWEEN SHIFT ANOVA RESULTS

CHOICE REACTION TIME			
Day	F (3,156) = 8.13	p=.001	Mean RT increased from d1 to d4 of the work week
Session	F (2,108) = 13.81	p=.001	Mean RT was generally quickest at the end of the workday
CHOICE REACTION TIME (ERRORS)			
Session	F (2,108) = 9.28	p=.001	Average RT errors increased from start to end of workday
Day x Session	F (6,309) = 3.98	p=.001	The increase in RT errors across sessions was more prominent at the end of the work week
DIGIT ADDITION (NUMBER ATTEMPTED)			
Day	F (3,155) = 3.81	p=.02	Average number of problems attempted increased across the work week
DIGIT ADDITION (NUMBER OF ERRORS)			
No Significant Findings			
GRAMMATICAL REASONING (RESPONSE TIME)			
Day	F (3,149) = 2.89	p=.04	Average response time generally declined from d1 to d4
Session	F (2,105) = 24.48	p=.001	Average response time was quickest at the end of the workday
GRAMMATICAL REASONING (ERRORS)			
No Significant Findings			

TABLE 2

WITHIN SHIFT ANOVA RESULTS

CHOICE REACTION TIME

8-H	Day	$F(4,106) = 7.59$	$p=.0001$	Mean RT increased from d1 to d5
	Session	$F(2,54) = 4.24$	$p=.02$	Mean RT was quickest at the end of the workday
	Day x Session	$F(8,211) = 3.17$	$p=.02$	Mean RT was quickest at the end of the workday for all except mid-shift
10-H	Day	$F(3,77) = 2.96$	$p=.04$	Mean RT increased from d1 to d4
	Session	$F(2,54) = 10.93$	$p=.001$	Mean RT was quickest at the end of the workday

CHOICE REACTION TIME (ERRORS)

8-H	Session	$F(2,54) = 11.14$	$p=.001$	Average RT errors were greatest at the end of the workday
	Day x Session	$F(8,211) = 3.70$	$p=.001$	Toward end of work week there were greater differences in the average RT errors from start to end of the workday
10-H	Session	$F(2,54) = 4.63$	$p=.02$	Average RT errors were greatest at the end of the workday

DIGIT ADDITION (NUMBER ATTEMPTED AND NUMBER OF ERRORS)

8-H/10-H No Significant Findings

GRAMMATICAL REASONING (RESPONSE TIME)

8-H	Session	$F(2,52) = 8.40$	$p=.001$	Average response time was quickest at the end of the workday
10-H	Session	$F(2,54) = 14.62$	$p=.001$	Average response time was quickest at the end of the workday

GRAMMATICAL REASONING (ERRORS)

8-H/10-H No Significant Findings

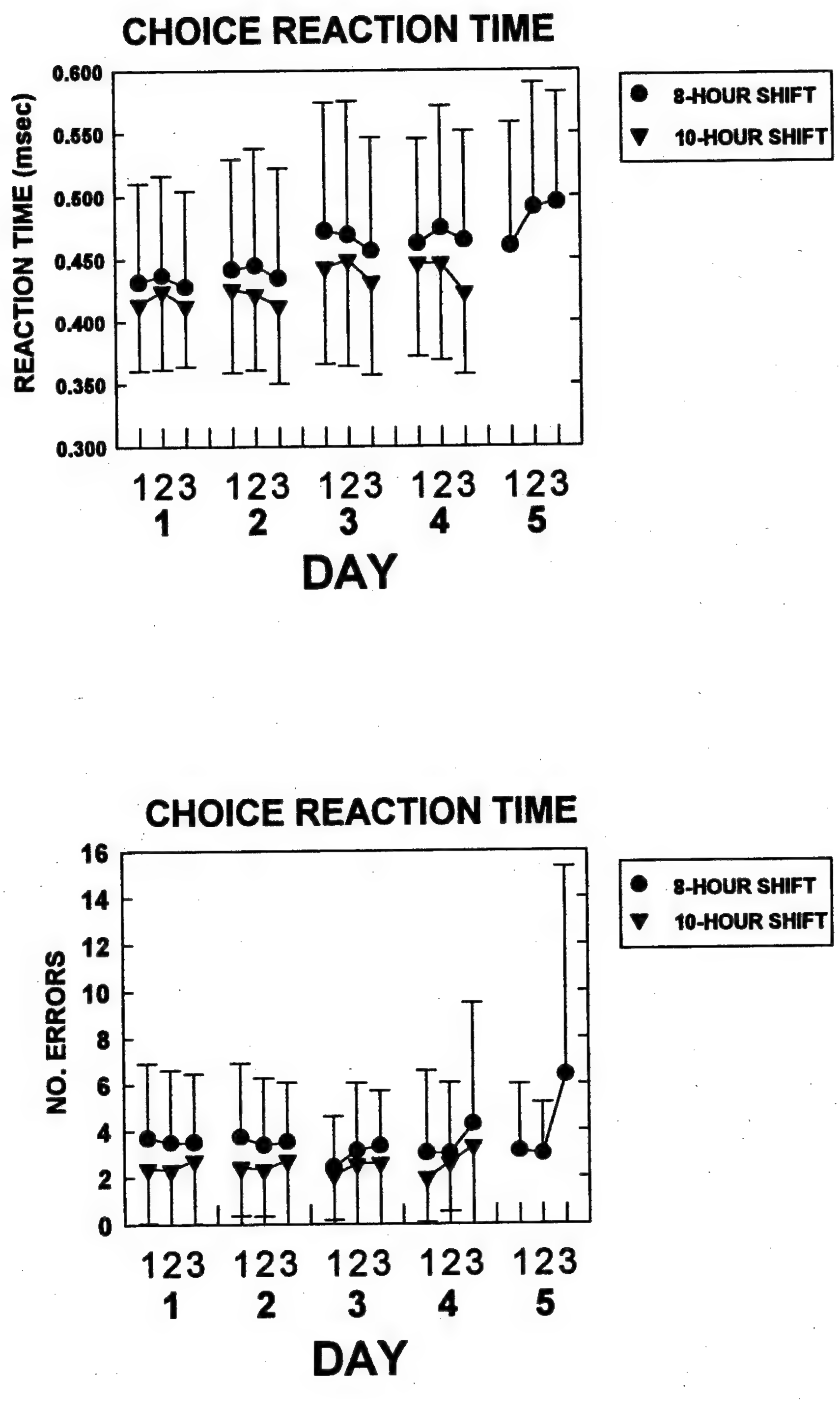
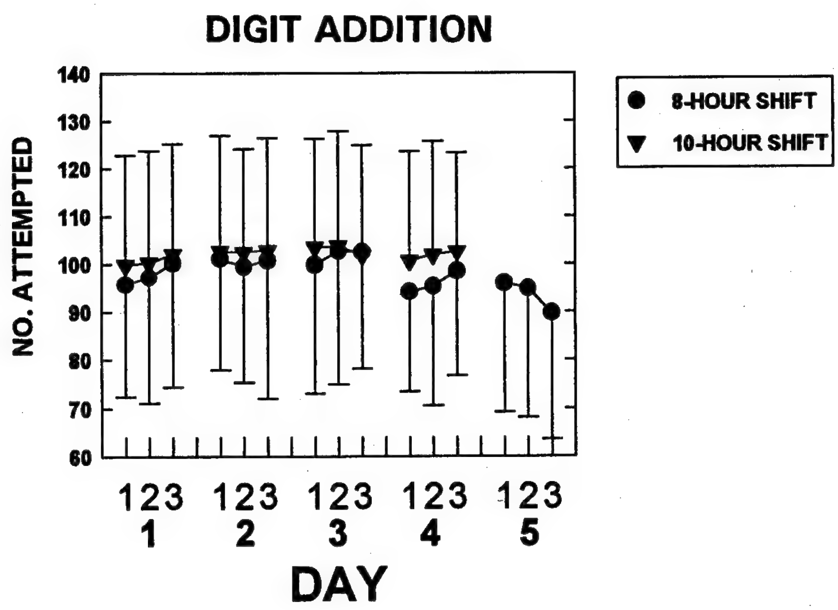


Figure 3. Means and standard deviations for the choice reaction times and average number of reaction time errors for each session of the day across the work week, for ATCSs working the 2 shift schedules.



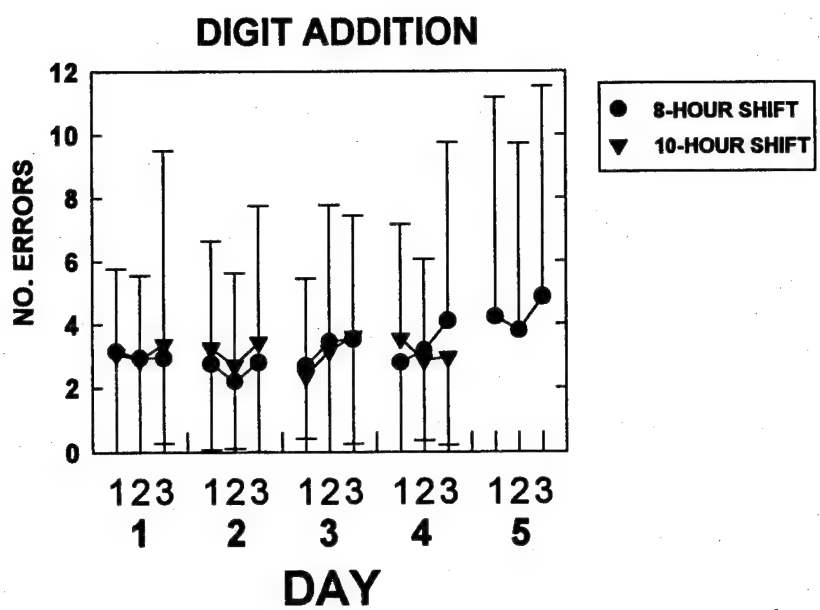
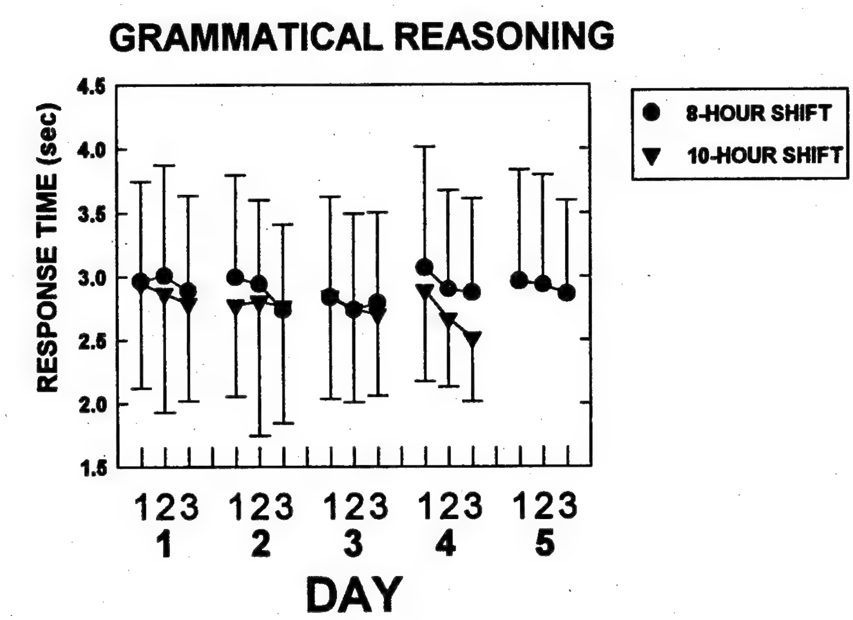


Figure 4. Means and standard deviations for the number of digit addition problems attempted and number of digit addition errors for each session of the day across the work week, for ATCSs working the 2 shift schedules.



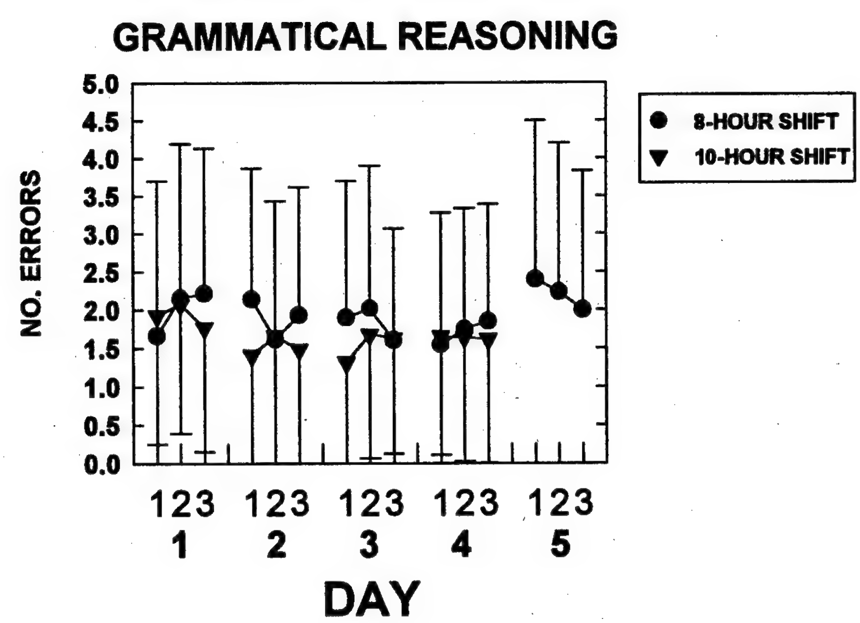


Figure 5. Means and standard deviations for response times on the grammatical reasoning test and average number of errors for each session of the day across the work week, for ATCSs working the 2 shift schedules.

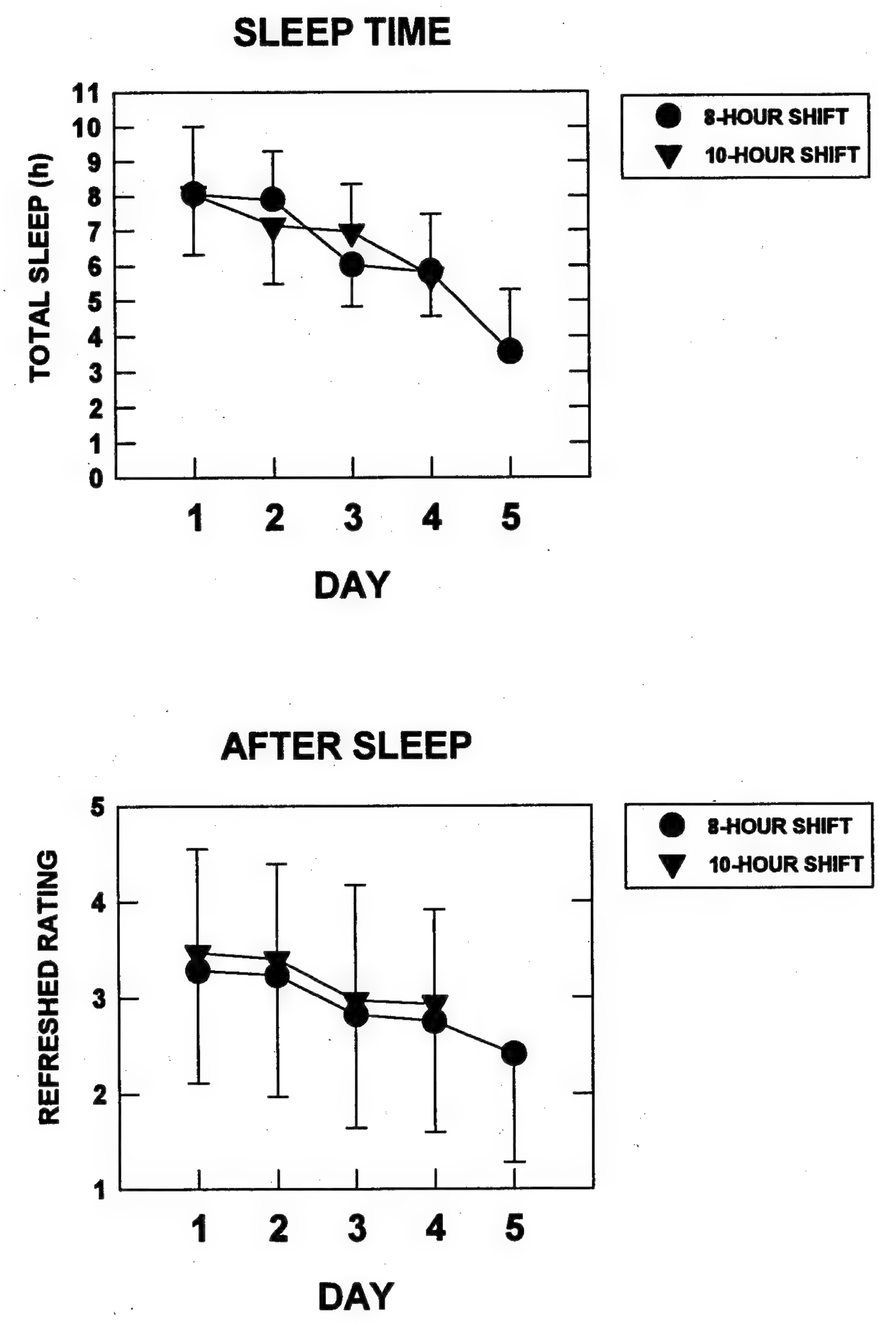


Figure 6. Means and standard deviations for total sleep time (in hours) and ratings of feeling "refreshed" following sleep for each day of the work week, for ATCSs working the 2 shift schedules.

Daily Sleep and Mood

Means for the total sleep time and ratings of feeling refreshed following sleeping are presented in Figure 6. The ANOVA for the sleep diary data yielded a significant effect for day $F(3,154) = 39.77$ $p=.001$ and shift by day $F(3,154) = 3.85$ $p=.02$. Both the 8-h and 10-h ATCSs exhibited a general decline in total sleep time from an average of 8.35-h on the evening prior to the first day of the work week to approximately 5.75-h on the evening prior to the fourth day of the work week. ATCSs on the 8-h shift exhibited the lowest average number of hours of sleep on the day prior to the mid shift (3.75 h).

Changes in subjective ratings of feeling refreshed following sleep corresponded to the changes noted in the amount of sleep. The ANOVA revealed a significant effect for day $F(3,155) = 9.38$ $p=.001$. Feeling refreshed declined from an average rating (based on a scale of 1 to 5) of 3.28 (8-h) or 3.46 (10-h) for the evening prior to the first day of the work week to 2.70 and 2.89 for the evening prior to the fourth day. The lowest rating was that of 2.50 for sleep that occurred during the day for ATCSs on the 8-h schedule prior to the mid-shift. Even though ratings for most of the other quality of sleep questions evidence a general decline from the first through final day of the work week, the overall differences were less prominent than those for "feeling refreshed following sleep."

Levels of positive and negative moods associated with work are presented in Figure 7. The between shift ANOVA for positive mood ratings yielded a significant effect for day $F(3,156) = 4.33$ $p=.006$ and day by session interaction $F(6,310) = 10.02$ $p=.001$. Positive ratings of mood remained relatively stable across the first 3 days of the work week for both groups. Ratings for ATCSs in the 8-h group declined for both days 4 and 5. On the afternoon shifts (days 1 and 2), positive mood ratings declined from start to close of the workday. For the morning shifts, ratings for the final 2 sessions of the workday were above those of the start of the day. Positive ratings for the mid-shift declined from start to close of the workday.

As is evident in Figure 7, changes in ratings of negative mood tend to mirror those noted for positive mood. Negative mood remained relatively stable across the first 4 days of the work week, with ATCSs on the 8-h mid-shift reporting the highest negative mood. While positive mood ratings decreased across sessions on day 1 and day 2, negative mood ratings increased in a corresponding manner. During the morning shifts, affect became generally more positive across the day. The largest change in mood across sessions occurred for ATCSs involved in the mid-shift, where positive mood declined and negative mood increased.

ATCSs involved in this study reported very few somatic complaints, an average of less than 2 complaints per individual per session. There was no evidence of any significant changes in the somatic complaints across either days of the work week or sessions.

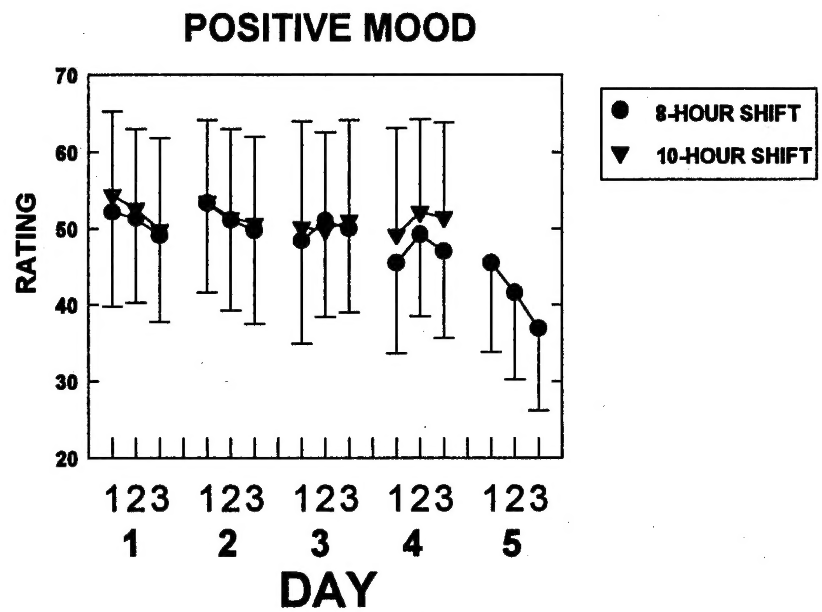
DISCUSSION

8-Hour Versus 10-Hour Comparisons

Our results suggest that ATC personnel working the 10-h shift schedule do not exhibit any evidence of lower performance on the NIOSH tests across workdays or within workdays than do ATCSs on an 8-h rotating schedule. Regardless of the task (reaction time, digit addition, or grammatical reasoning), none of the between-group differences in performance was statistically significant. This was true for both the reaction time and error measures. Any differences in test performance that were present tended to favor the 10-h ATCSs. This outcome is at contrast with the general findings from investigations concerning 12-h workdays. In comparing the 8-h and 12-h workdays, Rosa & Bonnet (1993), Rosa, Colligan & Lewis (1989), and Rosa (1991) all reported that performance on some tests from the NIOSH Fatigue Test Battery was significantly poorer for those employed on 12-h work schedules. However, studies involving the 12-h shift schedule included complete coverage of the 24-h workday, while the ATCSs involved in this 10-h study only covered the afternoon and morning shifts. At present ATCSs at this facility are not assigned to work a 10-h mid-shift.

Fatigue Test Battery Sensitivity

Results clearly demonstrated the sensitivity of the tests selected from the NIOSH fatigue test battery to alterations in alertness associated with working a rotating shift schedule. However, the tests included in this study did not appear to be equally sensitive to the effects of either the workday or sessions within



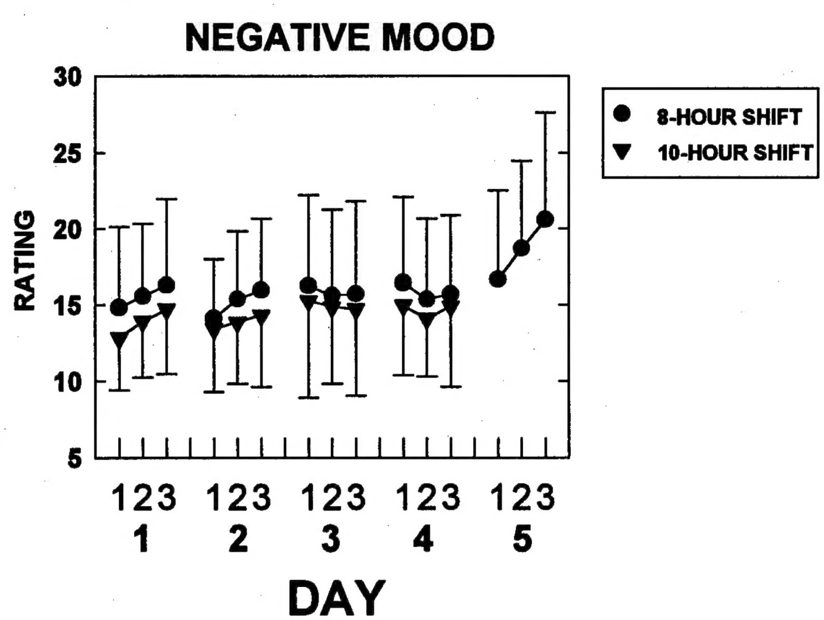


Figure 7. Means and standard deviations for positive and negative mood ratings for each session of the day across the work week, for ATCSs working the 2 shift schedules.

the workday. The choice reaction time measure appeared to be most sensitive to variations in test time and effects of the work week. The mid-shift, with the associated sleep loss and test times between 10 pm and the early morning hours clearly resulted in slower response times and greater errors. Changes in average values for ATCSs on the mid-shift were the only occasions where response times declined and errors increased from start to completion of the workday on each of the 3 tests. The extent to which these changes can be attributed to the sleep loss associated with this particular quick-rotating schedule, or to the effects of the circadian rhythm on performance, cannot be determined from this study.

The obtained alterations in the fatigue test battery performance measures reflect both the attentional demands of the specific tasks and the overall sensitivity of the component measures to fatigue, but were not necessarily reflected in changes in operational (job) performance. Operational tasks often involve much greater opportunity for analysis and response to critical situations than the tasks presented under these experimental conditions. However, the outcomes do reflect some general decrements in readiness of the human operator to respond that are associated with circadian rhythms and sleep loss resulting from a rotating shift schedule.

Sleep and Mood

The consistent decline in total sleep time from start to completion of the work week reported by the ATCSs who participated in this study appears to be closely related to the general pattern of changes in performance noted on the choice reaction time and digit addition components of the test battery. The effects were also readily observable in the self-reported positive and negative mood ratings and ratings of feeling "refreshed" following sleep. Shortened sleep times during the first 4 days were nearly identical for ATCSs on both the 8-h and 10-h shift schedules. From a high of approximately 8.3 hours on the evening prior to the first day of the work week, ATCSs reported progressively fewer hours of sleep each day to approximately 5.75 hours prior to day 4. For ATCSs on the 8-h schedule, the combination of the short turnaround and the need to sleep during the daylight hours, resulted in the shortest sleep time prior to the mid-shift on the final day of the work week (approximately 3.75 hours).

The number of hours of sleep reported by ATCSs working the 8-h shifts in this study are consistent with those reported in earlier studies by Melton and his colleagues in the 1970s (Melton, et al. 1971; 1973; 1975; and Saldivar, Hoffman, and Melton, 1977). Melton, et al. (1973) commented that the shorter sleep time for ATCSs prior to the mid-shift may be attributed, in part, to the tendency for some ATCSs to take only a brief nap prior to the mid-shift so that they will be able to sleep better during the morning following completion of the mid-shift. These findings are also consistent with outcomes from a recent assessment of controller sleep time on the 2-2-1 schedule at the Miami en route center (Cruz and Della Rocco, 1995), where ATCSs averaged 2.4 hours of sleep prior to the mid-shift. In contrast, controllers at the Miami International Flight Service Station reported slightly longer sleep periods (5.45- h) prior to the mid-shift (Melton, 1985). These outcomes suggest that individuals develop different strategies relative to the amount of sleep they obtain prior to the mid-shift.

Rapidly Rotating Shift Schedules

Despite some of the obvious advantages of rapidly-rotating shift schedules, where employees generally work 2 or fewer nights in succession, research concerning those schedules is not very extensive. Additionally, there is considerable variation in the actual working hours and nature of the proposed rapidly rotating schedules. Despite these facts, Wilkinson (1992), in a brief review of the outcomes from various types of shift schedules, concluded that fixed night systems are superior and should be implemented for night work. The effectiveness of this approach however, is dependent on the willingness of employees to remain on a "night" schedule even during their days off. In reply, Folkard (1992), argued that Wilkinson overestimated the problems associated with rapidly-rotating shift systems and that other aspects of shift systems should be taken into account when determining the best shift schedule. In a series of studies, Melton and his colleagues reported that while ATCSs on a 2-2-1 schedule obtain slightly less sleep across the work week than their colleagues on either a 5-day rotating or 5-day fixed-schedule, they did not differ significantly on most of the physiological and biochemical indices of stress (Melton, et al. 1971; 1973; 1975; and Melton, 1985), or the measures of mood and anxiety (Melton, et al. 1971; 1973; 1975). Melton (1985), however, reported that a group of ATCSs employed at the

Miami flight service station on the 2-2-1 shift exhibited higher levels of self-reported fatigue prior to the start of their work week than those on a 5-day fixed schedule.

ATCSs who favor the 2-2-1 schedule have consistently reported that this preference is based primarily on the longer number of hours off between work weeks, and that they are required to work only a single mid-shift. Another social factor associated with the 2-2-1 shift schedule is that a relatively normal amount of sleep and a relatively normal family schedule can be maintained during much of the work week. Ability to maintain a near normal pattern of sleep time is only seriously disrupted just prior to starting the mid-shift. Additionally, the timing of the change in shifts is such that the staff of ATCSs who handle the typical morning push of air traffic comes from the ATCSs who have just started their workday, rather than those who are completing the mid-shift.

While there is considerable variation in shift schedule preference among ATCSs, the 2-2-1 schedule has continued to be viewed positively by much of the ATC workforce. This is evidenced, in part, by its continued existence at most ATC facilities for more than 2 decades, as employees, union representatives, and management have conferred regarding the selection of a preferred shift schedule. Anecdotal comments from controllers and facility managers, however, suggest that the percentage of younger controllers preferring the 2-2-1 schedule is greater than that of older controllers. However, as part of an older survey of ATCS job attitudes, Smith (1973) determined that while there was a trend for the preference of the 2-2-1 schedule to diminish with age, it was still the most preferred schedule for older controllers. As the ATC workforce ages over the next decade, continued research will be needed to determine the extent to which older controllers may experience difficulties in coping with the 2-2-1 schedule, and to assess the effectiveness of alternative schedules and fatigue countermeasures that would reduce the negative consequences of working a rotating shift schedule. During the 2 decades of using the 2-2-1 shift schedule at ATC facilities across the U.S., controllers have provided anecdotal comments concerning difficulties associated with working a rotating shift schedule. However, there is little documented evidence of any significant negative impact on work performance, safety, or overall well-being.

REFERENCES

Baddeley, A. D. (1968). A 3-minute reasoning task based on grammatical transformation. *Psychonomic Science*, 10, 341-342.

Broach, D., & Aul, J. C. (1993). Analysis of the air traffic control specialist (ATCS) occupation using the Position Analysis Questionnaire (PAQ), unpublished manuscript.

Calvasina, E., & Boxx, R. (1975). Efficiency of workers on the four day work week. *Academy of Management Journal*, 18, 604-610.

Cruz, C. E., and Della Rocco, P. S. (1995). *Sleep patterns in Air Traffic Controllers working rapidly rotating shifts: A field study.* (DOT/FAA/AM-95/12), Washington, DC: Federal Aviation Administration.

Daniel, J., and Potasova, A. (1989). Oral temperature and performance in 8 h and 12 h shifts. *Ergonomics*, 32, 689-696.

Dunham, R. B., Pierce, J. L., & Castaneda, M. B. (1987). Alternative work schedules: Two quasi-experiments. *Personnel Psychology*, 40, 215-242.

Folkard, S. (1992). Is there a 'best compromise' shift system? *Ergonomics*, 35, 1453-1463.

Goodale, J. G. & Aagaard, A. K. (1975). Factors relating to varying reactions to the 4-day workweek. *Journal of Applied Psychology*, 60, 33-38.

Hartman, R. I., & Weaver, K. M. (1977). Four factors influencing conversion to four-day work week. *Human Resource Management*, 16, 24-27.

Hodge, B. J. & Tellier, R. D. (1975). Employee reactions to four-day work week. *California Management Review*, 18, 25-30.

Johnson, L. C. & Naitoh, P. (1974). *The operational consequences of sleep deprivation and sleep deficit.* AGARDograph AF-193, NATO Advisory Group for Aerospace Research and Development. London, England.

Kenny, M. T. (1974). Public employee attitudes toward the four-day work week. *Public Personnel Management*, 3, 159-161.

Latack, J. C., & Foster, L. W. (1985). Implementation of compressed work schedules: Participation and job redesign as critical factors for employee acceptance. *Personnel Psychology*, 38, 75-93.

Lewis, P. M., and Swaim, D. J. (1988). Effect of a 12-hour/day shift on performance, in E. W. Hagen (ed.) *Proceedings of the Fourth Institute of Electrical and Electronics Engineers Conference on Human Factors and Power Plants*. (pp. 513-516, Institute of Electrical and Electronics Engineers, New York, NY).

Melton, C. E. (1985). *Physiological responses to unvarying (steady) and 2-2-1 shifts: Miami International Flight Service Station.* (DOT/FAA/AM-85/2), Washington, DC: Federal Aviation Administration.

Melton, C. E., McKenzie, J. M., Polis, B. D., Funkhouser, G. E., and Iampietro, P. F. (1971). *Physiological responses in air traffic control personnel: O'Hare Tower*. (FAA/AM-71-2), Washington, DC: Federal Aviation Administration.

Melton, C. E., Smith, R. C., McKenzie, J. M., Saldivar, J. T., Hoffman, S. M., and Fowler, P. R. (1975). *Stress in air traffic controllers: Comparison of two air route traffic control centers on different shift rotation patterns*. (FAA/AM-75-7), Washington, DC: Federal Aviation Administration.

Melton, C. E., McKenzie, J. M., Smith, R. C., Polis, B. D., Higgins, E. A., Hoffman, S. M., Funkhouser, G. E., and Saldivar, J. T. (1973). *Physiological, biochemical, and psychological responses to air traffic control personnel: Comparisons of the 5-day and 2-2-1 shift rotation patterns*. (FAA/AM-73-22), Washington, DC: Federal Aviation Administration.

Mills, M. E., Arnold, B., and Wood, C. M. (1983). Core-12: A controlled study of the impact of 12-hour scheduling. *Nursing Research*, 32, 356-361.

Myers, J. L. (1979). *Fundamentals of experimental design*. (Allyn & Bacon, New York).

Peacock, B., Glube, R., Miller, M., and Clune, P. (1983). Police officers' responses to 8 and 12 hour shift schedules. *Ergonomics*, 26, 479-493.

Rosa, R. R. (1991). Performance, alertness and sleep after 3-5 years of 12 h shifts: a follow-up study. *Work & Stress*, 5, 107-116.

Rosa, R. R., & Bonnet, M. H. (1993). Performance and alertness on 8 h and 12 h rotating shifts at a natural gas utility. *Ergonomics*, 36, 1177-1193.

Rosa, R. R., and Colligan, M. J. (1988). Long workdays versus restdays: Assessing fatigue and alertness with a portable performance battery. *Human Factors*, 30, 305-317.

Rosa, R. R., Colligan, M. J., & Lewis, P. (1989). Extended workdays: Effects of 8-hour and 12-hour rotating shift schedules on performance, subjective alertness, sleep patterns, and psycho-social variables. *Work and Stress*, 3, 21-32.

Rosa, R. R., Wheeler, D. D., Warm, J. S., & Colligan, M. J. (1985). Extended workdays: Effects on performance and ratings of fatigue and alertness. *Behavior Methods, Instruments, and Computers*, 17, 6-15.

Saldivar, J. T., Hoffman, S. M., and Melton, C. E. (1977). *Sleep in air traffic controllers.* (FAA/AM-77-5), Washington, DC: Federal Aviation Administration.

Smith, R.C. (1973). *Job attitudes of air traffic controllers: A comparison of three air traffic control specialties*. (FAA/AM-73-2), Washington, DC: Federal Aviation Administration.

Steele, J. L., & Poor, R. (1970). Work and leisure: The reactions of people at 4-day firms, in R. Poor (Ed.), *4 Days, 40 Hours*. (pp. 105-122, Bursk and Poor, Cambridge, MA).

Volle, M., Brisson, G. R., Perusse, M., Tanaka, M., and Doyon, Y. (1979). Compressed work-week: Psychophysiological and physiological repercussions. *Ergonomics*, 22, 1001-1010.

Wheeler, M. E. (1970). Four days, forty hours: A case study. *California Management Review*, 17, 74-81.

Wilkinson, R. T. (1992). How fast should the night shift rotate? *Ergonomics*, 35, 1425-1446.

Williams, H. L., & Lubin, A. (1966). Speeded addition and sleep loss. *Journal of Experimental Psychology*, 73, 313-317.

☆U.S. GOVERNMENT PRINTING OFFICE: 1996 - 760-082/20040